## METHODS FOR REDUCED COST INSERTION OF VIDEO SUBWINDOWS INTO COMPRESSED VIDEO

### RELATED APPLICATIONS

The present application is a continuation-in-part of pending Provisional U.S. Patent Application S.N. 60/064,584 Filed October 20, 1997 which is titled "METHODS FOR REDUCED COST INSERTION OF VIDEO SUBWINDOWS INTO COMPRESSED VIDEO" and which is hereby expressly incorporated by reference.

### Field of the Invention

The present invention relates to image processing and, more particularly, to methods and apparatus of encoding digital images to facilitate the subsequent insertion of additional image data into previously encoded images and to methods and apparatus for inserting said additional image data.

### Background Of The Invention

There are known networked distribution systems for transmitting television programming whereby audio visual material is transmitted to a number of affiliated stations, each of which retransmits the programming to viewers' homes.  Fig. 1 illustrates a known system 10 for distributing analog video data, e.g., NTSC television signals. Within the known system, it is common practice

for the local affiliate station(s) 12, 14, 16 to electronically overlay a station logo (or other identification) or other local informational content over the network video signal, e.g., a corner portion of the

5   broadcast images.  In this manner, the video pictures that are presented to viewers, e.g., homes A-Z, include picture content that has been locally inserted.  In the case of analog television (e.g. NTSC) achieving the overlay of local content on the network signal can be

10  easily accomplished by electronically adding or switching video signals.

It appears that there is a significant desire to perform the local insertion of picture content within

15  emerging digital television networks.  In these digital networks, highly compressed digital video will be delivered to viewers' homes.  MPEG video coding will be used to accomplish this high degree of compression in many proposed digital television systems.

20

In order to obtain high compression efficiency, video compression techniques (e.g. MPEG) employ motion compensated prediction, whereby a region of pixels in the picture being compressed is coded by coding the

25  difference between the desired region and a region of pixels in a previously transmitted reference frame.  The term "motion compensated" refers to the fact that the position of the region of pixels within the reference frame can be spatially translated from the position of

30  the region being coded, in order to compensate for

motion, e.g., of the camera or in the scene, that occurred between the time that the two pictures were captured.

5          Fig. 2 illustrates the use of motion compensated prediction, in the form of motion vectors, to code a second image, Image 2 as a function of pixel values included in a reference frame, Image 1.  In Fig. 2, Image 1 and Image 2 both comprise a plurality of
10        segments including Segments A, B, and C.  Using the known motion compensated prediction techniques, a segment of Image 2 may be coded using a motion vector which refers to, e.g., a block of pixel values corresponding to a segment of Image 1 that is located outside the image
15        segment that is being coded.  By transmitting motion vectors as opposed to the actual pixel values, efficient motion compensated prediction coding of Image 2 is achieved.

20        Note that in the Fig. 2 prior art example Segment A of Image 2 is represented using motion vectors 23, 25 which reference Segments B and C of Image 1.  Motion vector 23 is an Image 2 motion vector which defines a portion of Image 2, Segment A as a
25        function of a portion of (Image 1, Segment C).  Motion Vector 25 is an Image 2 motion vector which defines a portion of (Image 2, Segment A) as a function of a portion of (Image 1, Segment B).

As discussed above, it is sometimes desirable to insert local image data into a previously encoded image. For example, a local broadcaster may want to insert a logo into Segment B of the image which is to be broadcast.

The use of motion compensated prediction makes it difficult for a local broadcaster to insert data into an encoded image by merely replacing encoded data blocks without running the risk of introducing errors into other frames which may rely on the image being modified as a reference frame. The difficulty of inserting new encoded image data in the place of previous encoded image data arises from the fact that the original coded blocks of subsequent coded images that are not part of the subset being replaced "assume" that the content of the replaced blocks is the original coded picture content. In such cases, any attempt to change the content of a subset of the blocks in the coded bitstream is likely to cause annoying prediction errors to propagate through the rest of the video, where blocks outside of the replaced subset were coded based on motion compensated predictions using picture content within the subset.

For example, if a logo was inserted into Segment B of Image 1 by substituting encoded data representing the logo for the encoded image data representing the original image content of Segment B of Image 1, a prediction error would result in Segment A of Image 2. Such a prediction error occurs because a

portion of the logo as opposed to the original image content of Segment B of Image 1 will now be incorporated into Segment A of Image 2 by virtue of the use of the motion vector 25.

5

In the absence of techniques for selectively replacing coded blocks of pixels in video bitstreams compressed with motion compensated prediction, there are two alternative approaches:

10

1)   Distribution of compressed video to affiliates (local stations) by forgoing the use of motion compensated prediction.  This would decrease compression efficiency so greatly that this approach would be

15   unacceptable for final transmission to viewers.

2)   Encoding or decoding and then re-encoding of a series of complete video images at the point of local transmission.  This approach removes motion vectors

20   generated by the original encoding and then generates an entirely new set of motion vectors based on the images into which data has been inserted.  This approach has the disadvantage of requiring the use of expensive video encoders at the local affiliate station capable of

25   encoding a series of complete images.  Also, this approach would generally require concatenated compression whereby video is first compressed for distribution to the affiliates, then fully decompressed and recompressed for final transmission, after local picture content is

30   inserted into the unencoded images generated by fully

decoding the originally encoded images. The application of concatenated compression generally results in picture degradation and when applied to complete images will, in many cases, result in final (home) picture quality which
5    is unacceptable.

Accordingly, there is presently a need for cost effective methods and apparatus which will support: 1) the transmission of encoded digital video data, 2) the
10   ability of local stations to insert sub-images and other local content into previously encoded video images; and 3) still provide an acceptable level of image quality to the final viewer of the encoded images, e.g., the home viewer of a video broadcast.

15

## Summary Of The Present Invention

The present invention comprises methods which permit bandwidth efficient video compression for
20   distribution to affiliates, and which allow the insertion of local picture content without the need for complete decompression and recompression. This is accomplished by the addition of a motion vector control module in the original video encoder which controls the selection of
25   motion vectors during the initial data compression (encoding) process. In accordance with the present invention, an encoder operator can define one or more non-overlapping subregions of the picture where local insertion is to be enabled. Alternatively, preselected
30   and predefined image subregions may be used at encoding

time. The motion vector control module determines the
minimum subset(s) of macroblocks in the picture that
encompass the defined subregion(s). During encoding, the
motion vector control module acts to guarantee that

5    motion vectors associated with macroblocks outside of
these subsets never result in the use of any pixels
contained within the subsets for constructing
predictions. Optionally, this module ensures that motion
vectors associated with macroblocks within a subset never

10    result in the use of pixels outside of the subset for
constructing predictions. Additionally, the encoder
contains a module which is capable of transmitting
information to affiliate stations regarding the size,
number and location of image subwindows or subregions

15    into which data may be inserted. The information
transmitted to a local station may include, e.g., the
number of subsets of macroblocks available for local
insertion, information identifying macroblocks belonging
to each subset, and information informing the local

20    station as to whether or not the optional motion vector
constraint described above was enforced.

        The present invention also involves an inserter
device, which would reside at the local affiliate station

25    that receives the information regarding the number and
placement of subsets of macroblocks that have been made
available for local insertion. An operator at the
affiliate station can specify the location and picture
content for local insertion. The inserter circuit parses

30    through the coded digital bitstream representing the

received encoded images, removes the data corresponding to the macroblocks that are affected by the desired local insertion, and replaces them with data corresponding to the desired local picture content. If it is desired that

5    the local picture content includes pixels from the original video, then some amount of decoding of the original video would be performed. In cases where the optional motion vector constraint, described above, is enforced, then only those bits corresponding to the

10   macroblocks need to be decoded. When the optional motion vector constraint were not enforced, then the inserter would decode some or all of the surrounding macroblocks in order to guarantee proper decoding of the pixels within the macroblocks that are to be affected by local

15   insertion.

Another approach involves the use of SNR scaleable coding. In this embodiment, the local picture content is added by the addition of an enhancement

20   bitstream, such as the SNR scaleable enhancement defined by MPEG-2. In addition to the local picture content, the SNR insertion device might optionally act to corrupt a small number of coded macroblocks of the network encoded bitstream, and encode the negative of the corruption

25   signal in the SNR enhancement layer. In this way a program provider might be able to encourage the purchase of receivers that are capable of the SNR scaleable decoding, and discourage the inactivation of this feature to avoid viewing the local picture content.

30

## Brief Description of the Drawings

Figure 1 illustrates a prior art system for generating and broadcasting analog NTSC television signals.

Figure 2 illustrates a prior art encoding technique involving the use of motion compensated prediction.

Figure 3 illustrates an encoding technique implemented in accordance with the present invention.

Figure 4 illustrates a digital video generation and transmission system implemented in accordance with one embodiment of the present invention.

Figure 5 illustrates a video data insertion circuit suitable for use in the system illustrated in Fig. 4.

## Detailed Description

As discussed above, the present invention relates to methods and apparatus of encoding digital images to facilitate the subsequent insertion of additional image data into previously encoded images. It also relates to methods and apparatus for inserting additional image data into data representing previously encoded images.

Figure 4 illustrates a digital video generation and transmission system 400 implemented in accordance

with one embodiment of the present invention.  The system
400 comprises a source of unencoded video data, e.g., a
video camera, an encoder 404 for encoding video data in
accordance with the present invention, a plurality of
5    local transmission stations 410, 412, 414 and a plurality
of  homes 440, 442, 444, 446, 448, 450.  The homes
include video decoders and display devices for viewing
the encoded video signals transmitted thereto from the
local stations 410, 412, 414.

10

          In the Fig. 4 embodiment, the video camera 402
generates unencoded video data in the form of an
unencoded digital video data stream.  Alternatively, the
video cameral 402 may output an analog video data stream.
15   The unencoded video data stream generated by the video
camera 402 is supplied to a motion compensated
encoder 405 included within the motion compensated
prediction encoder 404.  In the event that an analog
video signal is supplied to the encoder 404, it is
20   converted into a digital data stream using, e.g., an
analog to digital (A/D) converter, prior to performing an
encoding operation thereon.

          In accordance with the present invention,
25   images which are to be encoded are segmented into
different regions or sub-regions, representing different
image segments, for encoding purposes.  Information
regarding the image regions into which data may be
inserted by local stations is stored in a memory 407
30   which is included within the motion vector control

module 406. The stored insertion region information will normally identify one or more non-overlapping image subregions where local insertion is to be enabled. These subregions may be identified and input to the motion

5    vector control module 406 by an operator of the encoder 404 or pre-programmed into the memory 407. At encoding time, the use of pre-programmed sub-region information results in images being treated as having predefined image subregions which are to be treated

10   separately for motion compensated prediction purposes.

The motion compensated encoder 405 is responsible for encoding the video data received from the video camera 402 using motion compensated prediction

15   encoding techniques which involve the generation of motion vectors. The encoder 404 may be an MPEG-2 encoder which generates, as a function of a motion vector control signal, an MPEG-2 compliant bitstream from the video data input thereto. As will be discussed below, the motion

20   vector control signal controls which regions of a previous or subsequent image may be used as reference data when encoding various distinct regions, e.g., segments, of a current image.

25   The motion vector control module 406, includes computer software, logic circuitry or a combination of the two. The computer software may be stored in the memory 407 along with the sub-region information. The motion vector control module 406 controls the selection

and/or generation of motion vectors during compression, e.g., during encoding by the motion compensated encoder 405. The motion vector control module determines the minimum subset(s) of macroblocks in the picture that

5    encompass the image subregion(s) defined by the supplied or pre-selected sub-region information stored in the memory 407. For future reference, the subsets of macroblocks which define image subregions, e.g., image segments, where data may subsequently be inserted will be

10   termed "local insertion subsets", and the set of macroblocks that are not contained in any local insertion subsets will be termed the "main picture subset".

     The motion vector control module 406 checks to

15   insure that operator entered subregions do not overlap thereby ensuring that none of the resulting local insertion subsets are overlapping collections of macroblocks.

20   During encoding, the motion vector control module 407 controls the motion compensated encoder 404 to guarantee that motion vectors associated with macroblocks in the main picture subset do not use pixels included within the local insertion subsets for constructing

25   predictions. In one particular embodiment, the motion vector control module 406 also ensures that motion vectors associated with macroblocks within one or more of the local insertion subsets do not use pixels of preceding or adjacent images, outside of the particular

30   subset for constructing predictions. This requirement

that predictions generated to represent image segments
selected for data insertion be generated from a
corresponding image segment of a preceding or subsequent
image will be referred to herein as an optional motion
5      vector constraint since it is optional to the extent that
it is an additional constraint not included in the
previously discussed embodiment.

The encoder of the present invention also
10     includes a subregion information insertion module 407
which is responsible for supplying information specifying
the number of local insertion subsets, identifying the
macroblocks belonging to each local insertion subset,
and, for each local insertion subset, indicating whether
15     or not the optional motion vector constraint described
above was enforced.  In one embodiment, the encoder 405
encodes the information supplied by the subregion
information insertion module as auxiliary data which is
combined with the encoded video data prior to
20     transmission to the local stations 410, 412, 414.   Thus,
via the motion compensated encoder 405, the subregion
information insertion module 407 is capable of
transmitting information to affiliate stations 410, 412,
414 regarding the number of local insertion subsets,
25     identifying the macroblocks belonging to each local
insertion subset, and, indicating for each local
insertion subset, whether or not the optional motion
vector constraint described above was enforced.

The motion vector control module 406 can be embodied in a variety of ways. In one embodiment, it acts to limit the set of candidate motion vector values searched by the motion compensated encoder 405 during the

5    motion estimation process. In such an embodiment, the motion vector control module 406 constrains the motion estimation process performed by the encoder 405 to avoid consideration of motion vector values that would result in the use of disallowed pixels as prediction references.

10

In another embodiment, the motion vector control module 404 acts by determining whether the motion vector selected by the motion estimation process of the encoder 405 would result in the use of disallowed pixels

15    as prediction references and, when this is the case, substituting a motion vector value of zero to be used during actual encoding.

The motion compensated encoder of the present invention 404 generates a compressed video bitstream,

20    e.g., an encoded MPEG-2 compliant bitstream, which is supplied to a plurality of local stations 410, 412, 414 for distribution to end viewers. The local stations may be remotely located from the encoder 404 and coupled thereto, e.g., by satellite, cable or other high rate

25    digital communication medium.

Figure 3 illustrates the result of the encoding constraints of the present invention on images 1 and 2 previously illustrated as being encoded in the manner

30    illustrated in Fig. 2. Note that in the Fig. 3 example

of the present invention, segment B has been selected as
a segement to be used for local data insertion.  In
accordance with the present invention encoding of non-
segment B image segments using motion vectors which
5     reference segment B is prohibited.  Accordingly, in the
Fig. 3 example, segment A is encoded using a motion
vector 23 which references segment C but use of the
motion vector 25 which references segment B is prohibited
resulting in segment A being encoded without the use of
10    motion vector 25.

As discussed above, each local station may
choose to insert video data, e.g., advertisements or
emergency warnings, into the received video data prior to
15    distribution to the end viewers.  The video data 420,
422, 424 to be inserted at the local stations 410, 412,
414, respectively, is normally stored at the local
station, e.g., on magnetic tape.  In addition, or
alternatively, the local stations 410, 412, 414 may
20    insert data that is generated live, such as for emergency
messages.

In order to insert video data into the encoded
bitstream generated in accordance with the present
25    invention, each local station 410, 412, 414, includes a
video data insertion circuit 430, 432, 434, respectively.

Fig. 5 illustrates a video data insertion
circuit 452 suitable for use as any one of the video data

insertion circuits 430, 432, 434. The video data
insertion circuit 452 comprises a parser 454, a
decoder 456, an unencoded data combining circuit 458, an
encoder 460 and an encoded image data combining
5    circuit 462 coupled together as illustrated in Fig. 5.

The video data insertion circuit 452 receives
as its input compressed video, e.g., an encoded MPEG-2
bitstream, obtained form the encoder 404 and uncompressed
10   local video data to be inserted.

Since the information regarding the number and
placement of local insertion subsets and/or image
segments to be used for data insertion is included in the
15   received encoded data, the video data insertion circuit
452 has this information available to it. Such
information may be read from the received data by the
parser 454. In one embodiment where local insertion
subsets are preselected at the encoder, the preselected
20   insertion subsets are also programmed into the video data
insertion circuits 430, e.g., the parser 454. In such a
case, the insertion circuit 452 will be aware of the
supported local insertion subsets without the need for
obtaining the information from the received encoded
25   bitstream.

An operator of the insertion circuit 430, 432,
or 434 can select one or more of the local insertion
subsets, e.g., by supplying an insertion selection
30   information signal which identifies one or more insertion

subsets to be identified and used by the parser 454. The insertion selection information signal may, alternatively, identify the image location or locations into which the video data is to be inserted. The

5 insertion selection information signal may automatically be generated by a computer, e.g., using default selection values corresponding to an image area which is large enough to contain the image data to be inserted.

10 At the local station, the picture content, e.g., one or more logos, to be applied or inserted to each selected local insertion subset, is specified, e.g., by supplying the uncompressed local video data to be inserted to the video data insertion circuit 452.

15

In one embodiment the insertion device 452 uses the parser 454 to parse through the received coded bitstream. The parser 454 identifies and removes the data corresponding to the macroblocks in the selected

20 local insertion subsets. The received encoded data corresponding to the selected insertion subsets to be replaced or modified is supplied, along with additional data from the received bitstream required for accurate decoding, to the decoder 456.

25

In cases where the optional motion vector constraint, described above, was enforced for a given local insertion subset which is being used for data insertion resulting in the corresponding image segment

30 being encoded using motion vectors referencing the same

image segment of one or more different images, then only
those bits corresponding to the macroblocks in the local
insertion subset output by the parser 454 are decoded.
If the optional motion vector constraint is not enforced,
5    then the decoder 456 decodes some or all of the
surrounding macroblocks in order to ensure proper
decoding of the pixels within the macroblocks that are to
be affected by local insertion.

10       The decoder 456 operates to decode the received
data corresponding to the image portion to be replaced or
modified with the local data to be inserted.  The decoded
image data generated by the decoder 456 is used when the
desired effect of the locally inserted picture content is
15   a translucent overlay, or selectively transparent overlay
or otherwise involves the use of the received video.

         Thus, decoded video data corresponding to the
image area into which data is to be inserted is output by
20   the decoder 456.  This decoded video data is supplied to
a first input of the unencoded data combining circuit
458.  A second input of the unencoded data combining
circuit 458 receives, via an uncompressed local video
data input, the uncompressed local image data to be
25   inserted.

         In cases where the data to be inserted
corresponds to only a portion of the insertion segment or
is intended to be, e.g., a transparent overlay, requiring
30   some combination with the original image data, the

unencoded data combining circuit 458 is used to combine the unencoded video data to be inserted and the unencoded video data output by the decoder 456. When the locally supplied video data is intended to be inserted without

5      being combined with the original video data, the unencoded data combining circuit 458 is bypassed or disabled.

The unencoded video data output by the

10     combining circuit 458 is coupled to the input of the encoder 460. In the case where the local image data is not combined with the decoded data output by the decoder 456, the input to the encoder 460 will comprise the local data to be inserted. In cases of, e.g.,

15     overlays, the data input to the encoder 460 will be the result of a combination of received decoded data and the local data to be inserted.

The encoder 460 generates encoded image data

20     which is compliant with the encoding scheme used by the original encoder 404. Once generated the encoded data output by the encoder 460 is supplied to a first input of the encoded image data combining circuit 462.

25     In addition to receiving encoded image data to be inserted, the encoded image data combining circuit receives, from the parser 454, the received encoded image data less the data corresponding to the portion of the received image which was removed by the parser 454 to be

30     replaced by the data to be inserted.

The encoded image data combining circuit 462 combines the encoded data received from the parser 454 and the encoder 460 to generate an image or images which include the local video data which was to be inserted. The output of the combining circuit 462 serves as the output of the video data insertion circuit 452. It is this encoded video data, including the locally inserted image data, that is supplied to the homes coupled to the local station 410, 412, or 414 which inserted the data.

In some embodiments, the insertion circuit 452 may have to parse through the entire bitstream in order to extract and replace the bits corresponding to the selected local insertion subsets. The process of identifying, extracting and replacing data in the received encoded bitstream can, and in various embodiments is, simplified in various ways.

For example, in one particular embodiment, the initial encoder 404 includes a module incorporated into the encoder 405, that causes the coded macroblocks in the local insertion subsets to be immediately preceded by a synchronization code (e.g. a slice_start_code in the case of MPEG coding). Accordingly, in one embodiment a synchronization code is inserted by the encoder 405 immediately before bits representing an image segment intended to support data insertion. This simplifies the subsequent parsing operation performed by the insertion circuit 452 and allows for low complexity parsing of the

data to identify the location of the bits, e.g., by looking for the inserted header, corresponding to macroblocks of local insertion subsets.

5        In another exemplary embodiment the initial encoder 404 includes a module which provides information to affiliate stations as to the byte and bit locations of the coded macroblocks of the local insertion subsets. This allows the extraction and replacement for local

10       insertion to take place without the need for parsing bits of the main picture subset.

        In yet another exemplary embodiment SNR scaleable coding is used. In such an embodiment, the

15       local picture content is added by the addition of an enhancement bitstream, such as the SNR scaleable enhancement defined by MPEG-2. In addition to the local picture content, the SNR insertion device can optionally act to corrupt a small number of coded macroblocks of the

20       network encoded bitstream, and encode the negative of the corruption signal in the SNR enhancement layer.

        In this way a program provider could encourage the purchase of receivers that are capable of the SNR

25       scaleable decoding, and discourage the inactivation of this feature to avoid viewing the local picture content. This approach is applicable for systems where the ability to decode SNR scaleable bitstreams exist, e.g., in at least some of the homes where the encoded data is

30       ultimately decoded and displayed.

The present invention can be used to provide
cost effective, bandwidth efficient, network distribution
of compressed video with the ability to perform local
5    insertion of picture content without complete
recompression.